

Citation/Reference	<p>S Guha, V Van Belle, C Bottomley, J Preisler, V Vathanan, A Sayasneh, C Stalder, D Timmerman, T Bourne, (2013)</p> <p>External validation of models and simple scoring systems to predict miscarriage in intrauterine pregnancies of uncertain viability</p> <p>Human Reproduction, 28 (11), 2905-11</p>
Archived version	Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher
Published version	http://humrep.oxfordjournals.org/content/28/11/2905.long
Journal homepage	http://humrep.oxfordjournals.org/
Author contact	<p>vanya.vanbelle@esat.kuleuven.be</p> <p>+ 32 (0)16 32.10.65</p>
IR	https://lirias.kuleuven.be/handle/123456789/417710

External validation of models and simple scoring systems to predict miscarriage in intrauterine pregnancies of uncertain viability

Running title:

Predicting miscarriage in pregnancies of uncertain viability

S Guha^{1,2}, V Van Belle^{3,4}, C Bottomley⁵, J Preisler², V Vathanan⁵, A Sayasneh², C Stalder², D Timmerman⁶, T Bourne^{2,6}

1. Hillingdon Hospital NHS Trust, Pield Heath Road, Middlesex, UB83NN, UK
2. Queen Charlotte's & Chelsea Hospital, Imperial College, Du Cane Road, London, W12 0HS, UK
3. ESAT-SCD / iMinds-KU Leuven Future Health Department, KU Leuven, Leuven, Belgium
4. Department of Mathematics and Statistics, Liverpool John Moores University, Liverpool, UK
5. Department of Obstetrics and Gynecology, Chelsea and Westminster NHS Trust, Fulham Road, London, UK
6. Department of Obstetrics and Gynecology, University Hospitals Leuven, Campus Gasthuisberg, KU Leuven, Belgium

Corresponding author:

Dr Sharmistha Guha

3 Rosebank Cottages

London Road

Loudwater

HP10 9TA

Email: sharmistha.guha@gmail.com

Mob: 07963691608

External validation of models and simple scoring systems to predict miscarriage in intrauterine pregnancies of uncertain viability

ABSTRACT

Study Question:

Does a logistic regression model and scoring system to predict viability of an intrauterine pregnancy of uncertain viability (PUV) perform as well in an independent patient group as the original patient group?

Summary Answer:

The model and scoring system showed good performance on external validation confirming their value for the prediction of miscarriage/viability in PUV patients up to 11 to 14 weeks gestation. A new model and scoring system without gestational age showed reasonable test performance and form a valuable alternative when patients are not able to report their LMP.

What is known already:

Several individual ultrasound and demographic factors have been described as predictors for miscarriage. A logistic regression model and simple scoring system using basic clinical and ultrasound features such as maternal age, bleeding score, mean gestational sac diameter (MSD) and presence or absence of yolk sac have been developed to allow patient specific prediction of viability of PUV beyond the first trimester.

Study Design, size, duration:

Prospective observational external validation cohort study in two inner city early pregnancy assessment units over a period of 18 months.

Participants/materials, setting, methods:

All consecutive women with a PUV were recruited. Ultrasound (mean sac diameter and presence of yolk sac) and demographic variables (maternal age, bleeding score and gestational age) were noted. The outcome measure was first trimester (11 -14 week) viability. Women with unknown first trimester outcome were excluded. Receiver operating characteristics (ROC) curves and calibration plots were constructed. Test performance was compared to the original development dataset. A new model and scoring system, which did not include gestational age, was built and evaluated.

Main results and the role of chance:

575 women were recruited. Outcome was known for 89.2%. The model could only be validated in 400 patients, due to missing values in model variables and output. The model predicted viability with an area under the ROC Curve (AUC) of 0.845 (95% CI, 0.806-0.884) compared to 0.774 (95% CI, 0.701-0.848) in the original study. The AUC for the scoring system was 0.832 (95% CI, 0.792-0.872) compared to 0.771 (95% CI, 0.698–0.844) from the original study dataset. The new model and the scoring system, excluding gestational age, could be evaluated

on 503 patients and resulted in an AUC of 0.801 (95% CI, 0.762-0.841) for the model and 0.773 (95% CI, 0.733-0.812) for the scoring system.

Limitations, reasons for caution:

Approximately 22% patients could not be validated due to missing variables and 11% patients did not have their first trimester outcome.

Wider implications of the findings:

Both the model and the scoring system showed excellent performance on external validation confirming their generalisability and utility in prediction of viability beyond the first trimester clinical practice.

Study funding/Competing interest(s):

TB is supported by the Imperial Healthcare NHS Trust NIHR Biomedical Research Centre. This research is supported by Research Council KUL GOA MaNet, iMinds 2012, Belgian Federal Science Policy Office IUAP P719. VVB is a postdoctoral fellow of the Research Foundation – Flanders (FWO).

Key words:

Gestational sac, miscarriage, prediction models, pregnancy of uncertain viability, ultrasound.

INTRODUCTION

In a systematic review of the accuracy of first trimester ultrasound for the diagnosis of early embryonic demise, Jevé et al found that there was a paucity of high quality prospective data on which to base guidance (1). Three further recent publications suggested that criteria used to define miscarriage based on transvaginal scan (TVS) measurements of gestation sac and embryo size were unreliable (2,3,4). This new information led to the Royal College of Obstetricians and Gynaecologists (RCOG) in the United Kingdom amending their guidance such that on a single TVS a miscarriage can only be diagnosed when an empty gestation sac is > 25mm in mean diameter or when the crown-rump length (CRL) measurement for an embryo with no visible heartbeat is > 7.0 mm. The guidance emphasised the need for repeating

ultrasound scans at an interval in order to definitively comment on viability (5). At a consensus meeting of the Society of Radiologists in Ultrasound (SRU) in 2012 it was concluded that similar guidance should be adopted in the United States. Whilst this more conservative approach to the diagnosis of miscarriage is welcomed, it is likely to be associated with more pregnancies being classified as being of uncertain viability. The introduction of sensitive home pregnancy tests means women may have a positive pregnancy test even before they have missed a period. The result of these developments means women attending for TVS in early pregnancy are more likely to have an uncertain outcome and be asked to undergo a repeat examination at an interval of 7 -14 days.

We know that early pregnancy complications may lead to significant psychological morbidity (6). The uncertainty of waiting for repeat examinations may compound what is already a anxious situation for couples, furthermore there is often a reluctance to accept that definitive answers cannot always be given at a single visit. It is helpful to counsel women about the probable outcome of the pregnancy so they are better prepared for the likely findings when they return to the clinic.

Anxiety also arises when a pregnancy is found to be smaller than expected for the menstrual dates given by the patient. This may be an innocent finding and relate to variations in both the timing of ovulation and implantation (7). However it has been appreciated that a difference in embryonic size from that expected may reflect a higher risk of miscarriage or aneuploidy (8). For other women the discrepancy between expected and observed embryo measurements may be because they have irregular periods or do not accurately remember the date of their last menstrual period (LMP). Giving appropriate advice to women about the possible outcome of their pregnancy in these circumstances is important to avoid unrealistic expectations.

To address some of these issues we have previously described a mathematical model (M_o) and simple scoring system (SS_o) to predict the outcome of a pregnancy when it is classified as an intrauterine pregnancy of uncertain viability (PUV) (9). Both the model (M_o) and the scoring system (SS_o) provide an individualized probability of pregnancy viability depending on maternal age, the amount of vaginal bleeding, gestational age, the mean gestational sac diameter and visibility of a yolk sac. This model (M_o) and scoring system (SS_o) were developed on data from one hospital unit, and showed very good prediction of viability on internal validation.

However before introducing a test into routine clinical practice it is necessary to demonstrate a good performance in different settings and populations.

In the present study we carried out an external validation to test the performance of both the model (M_o) and scoring system (SS_o). The two units are inner London hospitals with large clinical throughputs and ethnically diverse populations. We also approached the problem of how to include women who are uncertain about the date of their LMP in the prediction model. We therefore used the training set from the original paper (9) to develop a new model (M_n) and scoring system (SS_n) that did not take gestational age into account. This model and the scoring system were then tested on both the original and external validation data.

MATERIALS AND METHODS

This was an 18 months multicentre prospective cohort study (January 2011 – July 2012) in the early pregnancy assessment units of Chelsea and Westminster Hospital (C&W) and Queen Charlottes and Chelsea Hospital, Imperial College, London (QCCH). Both units take referrals from primary care physicians, emergency units and other departments in the hospital. There is no minimum gestational age criteria for attendance.

Inclusion criteria for the study were: all women classified as having an intrauterine pregnancy of uncertain viability (PUV). A PUV was defined using the following criteria on the basis of a TVS:

The visualisation of a single intrauterine gestational sac with a mean gestation sac diameter (MSD) of $\leq 20\text{mm}$ at C&W (prior to September 2011) and $\leq 25\text{mm}$ at QCCH, with or without a visible yolk sac, or a single intrauterine gestational sac with a visible embryo of $\leq 6\text{mm}$ at C&W (prior to September 2011) and $\leq 7\text{mm}$ at QCCH but no embryonic heart beat. Exclusion criteria were women with multiple pregnancies and those who underwent termination of pregnancy.

Prior to the consultation, the women completed a written questionnaire which included the date of their LMP, previous obstetric history and pregnancy symptoms. This information was confirmed by the examining sonographer. The amount of vaginal bleeding was estimated using a pictorial bleeding assessment chart (PBAC) (10). The bleeding score ranged from 0 (no bleeding) to 4 (bleeding with clots).

All women underwent a TVS performed by an appropriately trained examiner (a nurse specialist, gynecologist or sonographer) using either Voluson® E8 Expert (GE healthcare, Wisconsin, USA) or Medison Accuvix® XG (Samsung Medison, Seoul, South Korea) ultrasound machine using a tight curvilinear transvaginal probe operating at a frequency of 3.5 – 9.3 MHz. Structured assessment consisted of measurements of the gestational sac and yolk sac in three orthogonal planes and measurement of the crown-rump length of the embryo if present. All images were stored in the Picture Archiving and Communication System (PACS) or as hard copies in the patient's casenotes. The measurements were reported using commercially available ultrasound databases (Radcentre (iSOFT – IBA healthgroup company) and Astraia (astraia software gmbh, Munich, Germany)).

All women included were scheduled to have a repeat TVS examination after 7 to 14 days. The outcome of interest was viability at the routine end of first trimester scan (11 to 14 weeks gestation).

Statistical Analysis

In our validation study, only the five variables included in the previously developed model were recorded. In the original model, univariable and multivariable logistic regression analysis were performed to establish relationship between first trimester viability and a number variables like maternal age, ethnicity, obstetric history, gestational age by LMP, abdominal pain, pain score, presence of vaginal bleeding, bleeding score, mean sac diameter, presence of yolk sac, mean yolk sac diameter and subchorionic hematoma. Multivariable models were developed with the significant variables, using a stepwise approach (9). These were maternal age, gestational age (calculated from LMP), PBAC, mean gestation sac diameter and presence or absence of a yolk sac. For both the mathematical model (M_o) and simple scoring system (SS_o), a receiving operating characteristics (ROC) curve was constructed to describe the relationship between the sensitivity and the false-positive rate ($1 - \text{specificity}$) to predict ongoing viability. Calibration plots, plotting the observed versus the predicted probability of viability for both the mathematical model and the scoring system, were constructed.

As the model and the scoring system could not be used in women with unknown LMP, a new mathematical model (M_n) and scoring system (SS_n) were developed on the training set of the original cohort. This model and scoring system were dependent on four variables which were maternal age, PBAC, mean sac diameter and presence or absence of a yolk sac. The new model (M_n) and the scoring system (SS_n) were applied to the original test set, and further tested on this new external validation dataset.

RESULTS

During this study period 575 consecutive women with a PUV were recruited. The outcome was known in 513 (89.2%) pregnancies. 260/513 (50.7%) pregnancies were viable and 253/513 (49.3%) non-viable at the end of the first trimester. Analysis was performed on all 400 patients for whom all covariates and the outcome were known. Most exclusions (n=103) were due to unknown LMP. The bleeding score and presence of the yolk sac was unknown for 9 and 2 pregnancies, respectively wherein 1 patient had both variables missing. Of these 400 pregnancies, 200 (50%) were viable at the end of first trimester. The data collection method is shown in figure 1. Patients who miscarried either had heavy bleeding in the interim with no pregnancy visualised on the follow up scan at 7 to 14 days, had a non-viable pregnancy diagnosed at a 7-14 day follow up scan or had a viable pregnancy at the follow up scan but subsequently miscarried before the end of the first trimester.

The model (M_o) was defined as

$$z = 9 - 0.27 \times \text{maternal age} - 0.4 \times \text{PBAC score} + 1.3 \times (\text{gestational age} < 42 \text{ days}) \\ - 0.09 \times (|\text{MSD} - 7|) + 1.3 \times (\text{yolk sac present}),$$

where maternal age is taken to be 35 when the age is less than 35. The estimated chance of a viable pregnancy is then obtained as $\exp(z) / (1 + \exp(z))$. On the new external validation set, this model obtained an area under the ROC curve (AUC) of 0.845 (95% CI 0.806-0.884). The scoring system (SS_o) obtained an AUC of 0.832 (95% CI 0.792-0.872) in the same set. The ROC curves of both the model and the scoring system are illustrated in Figure 2.

Since patients are often unable to report an LMP, and the original model is not able to obtain a risk estimate for these patients, a new model (M_n) and scoring system (SS_n) that did not include information on LMP, were built using the training set from (9). The resulting model is defined as

$$z = 10.87 - 0.29 \times \text{maternal age} - 1.03 \times \text{PBAC score} - 0.13 \times (|\text{MSD} - 7|) \\ + 0.77 \times (\text{yolk sac present}) + 0.91 \times \text{PBAC score} \times (\text{yolk sac present}),$$

where maternal age was again taken to be 35 when age is less than 35. The chance was obtained as before. This model was validated on 503 patients with known outcome and covariates of the external validation set. The model obtained an AUC of 0.801 (95% CI=0.765-0.841) on the

new external validation set. Since the information on LMP is not taken into account, the performance is less than for the original model.

To facilitate application of the model M_n , a scoring system was derived from the new model as in the original study (9,11). Table 1 summarizes the new score system SS_n . The score system obtained an AUC of 0.773 (95% CI=0.733-0.812) on the new external validation data. The ROC curves for the new model and scoring system are given in Figure 3. Calibration plots (not shown) illustrate that the models and score systems are well calibrated, indicating that the chance predicted by the model corresponds to the chance observed in the dataset.

For clinical practice, we propose to use the original model M_o or scoring system SS_o whenever the patient reports an LMP, and to use the new model M_n or scoring system SS_n otherwise. The color based representation in Figure 4 can be used for this purpose. This can also be applied to mobile phone and other softwares.

DISCUSSION

We have shown that a mathematical model developed to predict the outcome of pregnancies of initially unknown viability maintains its performance when subjected to external validation. We have also demonstrated that a new model that does not require knowledge of LMP as a variable has reasonable test performance. This is important as our study showed that 20% of women with a PUV can not accurately recall the date of their LMP.

Several tools have been described in the literature for predicting the outcome of pregnancies of uncertain viability, however these usually require both information derived from ultrasonography and biochemical parameters. Bignardi et al found that an hCG ratio > 2.0 was predictive of a viable pregnancy at 11-14 weeks (13). However this study required two blood tests and was based on a population of women initially classified with pregnancies of unknown location. Another logistic regression model was originally published by Elson et al in 2003 (14). This model was dependent on maternal demographics and ultrasound features but also required serum progesterone measurement. This model showed reasonable performance on temporal validation with an AUC of 0.85 compared to 0.97 which was the AUC in the original study (15). However this study took six years to recruit because over 90% of eligible women refused the test, possibly because of the necessity for blood tests. In routine clinical practice, bHCG is not performed for women who are diagnosed as PUV at the initial TVS. Therefore,

our study was based on simple indicators that would be normally ascertained as part of routine clinical practice.

An advantage of the mathematical model M_o and scoring system SS_o (9) validated in this paper, is that it can provide women with an individualised probability of the viability of their pregnancy at the end of the first trimester, only using demographic information, symptoms and TVS findings. Furthermore the risk of miscarriage can be given immediately following their examination. For example, a 35 year old woman with a bleeding score of 2 at 48 days gestation with an mean gestational sac diameter of 14mm and absent yolk sac has an 18% probability of viability at the end of first trimester using the original scoring system SS_o , and 13.2% using the new model SS_n that did not take gestational age into account.

The new statistical model (M_n) and scoring system (SS_n) which do not require LMP as a variable did not perform as well as the original model (M_o) and scoring system (SS_n). However, they showed reasonable test performance which was retained on external validation, and so give us a tool which we can use in the significant minority of women with uncertain dates.

In comparison to the original study (9), the external validation dataset showed a similar miscarriage rate (50% versus 50.7%). Approximately 39.8% women presented with symptoms of bleeding compared to 35.5% in the original population. Both the development and external validation studies took place in ethnically diverse busy inner city hospitals. The scoring system is easy to use and does not require a computer for use (see Figure 4). The model and scoring system can be integrated into any of the commercially available ultrasound reporting programs or into a smartphone or tablet application, and so their introduction into clinical practice should be straightforward.

At present, a multicenter prospective study is ongoing to better define the cut off levels for measurements of mean gestational sac diameter and CRL that can be used to diagnose miscarriage (Diagnosis of Miscarriage (DOM) Study). It is likely that this would support recommendations that more women are asked to return for a repeat scan to definitively check viability. A reliable estimated probability of the likely outcome of their pregnancy would be helpful when counselling these women. In the event of a high likelihood of a viable pregnancy, women could be reassured that they do not need close follow up. When the predicted outcome

is a non-viable pregnancy, women could be warned regarding possible heavy bleeding and offered a repeat TVS assessment after an interval of 7 to 14 days.

A very important issue in early pregnancy care is managing the realistic expectations and anxieties of women who present with bleeding. Our experience is that when given a choice, women prefer to be given an immediate individualised probability of the likely outcome of their pregnancies rather than being given the standard advice that there is a background 50% chance of miscarriage for all women with a PUV. A similar study using a predictive model and scoring system to predict the chance of viability in all women seen in an early pregnancy assessment unit, including those with an embryonic heart beat at initial TVS is also being validated at other units to assess its generalisability and patient acceptability (17).

Women with bleeding in early pregnancy are invariably anxious about the outcome of their pregnancy. Though our outcome measure is the first trimester viability it is a known fact that the risk of miscarriage in the 2nd trimester, after a normal dating scan at 11 – 14 weeks, is very low (16). We have shown that a prediction model or scoring system to predict viability works in different clinical settings. These prediction tools can be introduced into clinical practice to both counsel women about the likely outcome of their pregnancy, and offer guidance on appropriate follow up arrangements based on the prognosis for that specific pregnancy.

AUTHOR'S ROLES

CB conceptualised the idea of the article. SG and CB designed the study. SG, VV, JP collected the data. VVB performed the statistical analysis. SG drafted the manuscript. VVB, CB, AS, CS and DT contributed to writing of the article. TB had the original idea for the article, participated in data analysis, decision making and supervised writing of the article.

FUNDING:

Tom Bourne is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at Imperial College Healthcare NHS Trust and Imperial College London. The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

This research is supported by Research Council KUL GOA MaNet, iMinds 2012, Belgian Federal Science Policy Office IUAP P719. VVB is a postdoctoral fellow of the Research Foundation – Flanders (FWO).

CONFLICT OF INTEREST:

None

REFERENCES:

1. Jeve Y, Rana R, Bhide A, Thangaratinam S. Accuracy of first-trimester ultrasound in the diagnosis of early embryonic demise: a systematic review. *Ultrasound Obstet Gynecol*, 2011 Nov;38(5):489-96
2. Abdallah Y, Daemen A, Kirk E, Pexsters A, Naji O, Stalder C, Gould D, Ahmed S, Guha S, Syed S, Bottomley C, Timmerman D, Bourne T. Limitations of current definitions of miscarriage using mean gestational sac diameter and crown-rump length measurements: a multicenter observational study. *Ultrasound Obstet Gynecol*, 2011 Nov; 38(5):497-502.
3. Pexsters A, Luts J, Van Schoubroeck D, Bottomley C, Van Calster B, Van Huffel S, Abdallah Y, D'Hooghe T, Lees C, Timmerman D, Bourne T. Clinical implications of intra- and interobserver reproducibility of transvaginal sonographic measurement of gestational sac and crown-rump length at 6-9 weeks' gestation. *Ultrasound Obstet Gynecol*, 2011 Nov;38(5):510-5.
4. Abdallah Y, Daemen A, Guha S, Syed S, Naji O, Pexsters A, Kirk E, Stalder C, Gould D, Ahmed S, Bottomley C, Timmerman D, Bourne T. Gestational sac and embryonic growth are not useful as criteria to define miscarriage: a multicenter observational study. *Ultrasound Obstet Gynecol*, 2011 Nov; 38(5) : 503-9.
5. RCOG. The Management of early pregnancy loss. Greentop Guideline No. 25. RCOG Press. London 2006 and Addendum to Greentop Guideline No. 25. RCOG Press. London, 2011
6. Lok IH, Neugebauer R. Psychological morbidity following miscarriage. *Best Pract Res Clin Obstet Gynaecol*, 2007 Apr;21(2):229-47.
7. Mahendru A, Daemen A, Everett T, Wilkinson I, McEniery C, Abdallah Y, Timmerman D, Bourne T, Lees C. Impact of ovulation and implantation timing on

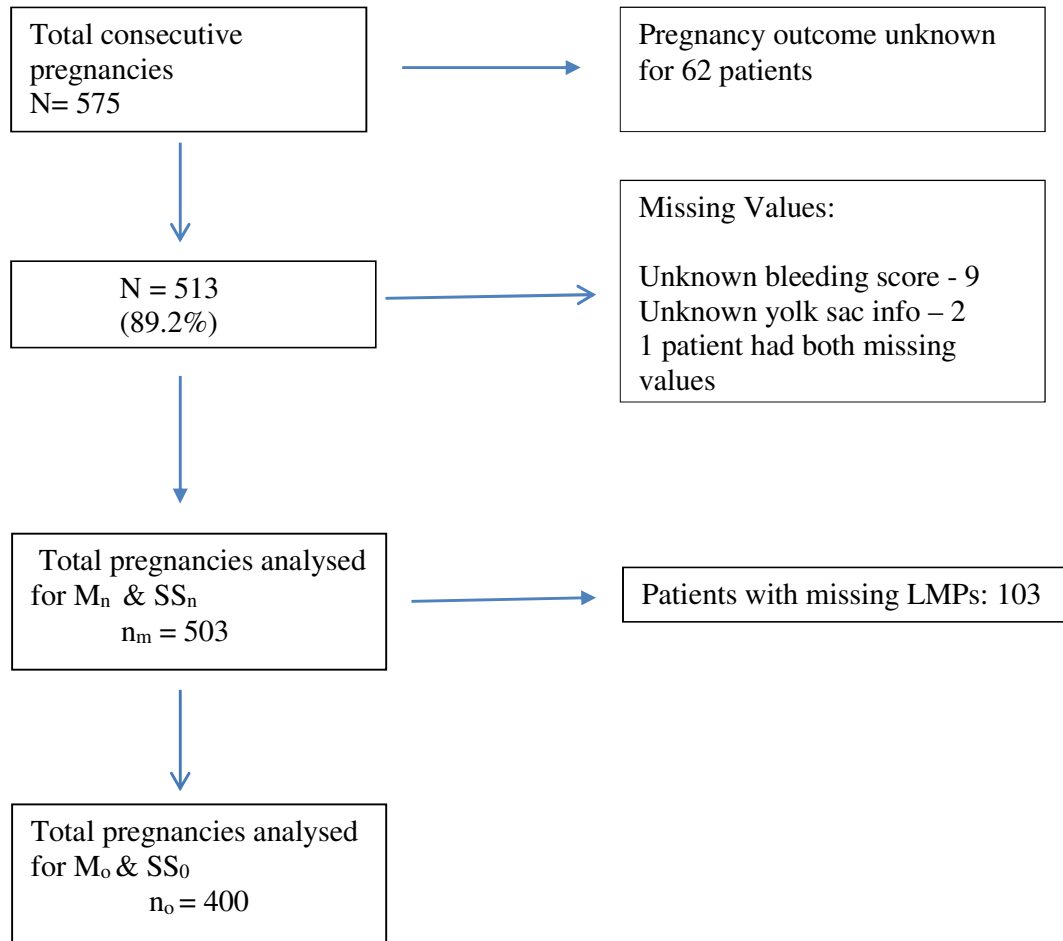
- first trimester crown-rump length and gestational age. *Ultrasound Obstet Gynecol*, 2012 Aug 1. doi: 10.1002/uog.12277. [Epub ahead of print]
8. Mukri F, Bourne T, Bottomley C, Schoeb C, Kirk E, Papageorghiou AT. Evidence of early first-trimester growth restriction in pregnancies that subsequently end in miscarriage. *BJOG*, 2008 Sep;115(10):1273-8.
 9. Bottomley C, Van Belle V, Pexsters A, Papageorghiou AT, Mukri F, Kirk E, Van Huffel S, Timmerman D, Bourne T. A model and scoring system to predict outcome of intrauterine pregnancies of uncertain viability. *Ultrasound Obstet Gynecol*, 2011 May; 37(5):588-95.
 10. Higham JM, O'Brien PM, Shaw RW. Assessment of menstrual blood loss using a pictorial chart. *Br J Obstet Gynaecol*, 1990; 97: 734–739.
 11. Sullivan LM, Massaro JM, D'Agostino RB. Presentation of multivariate data for clinical use: The Framingham Study risk score functions. *Stat Med*, 2004; **23**: 1631–1660.
 12. Van Belle VM, Van Calster B, Timmerman D, Bourne T, Bottomley C, Valentin L, et al. A mathematical model for interpretable clinical decision support with applications in gynecology. *PLoS ONE*, 2012;7(3):e34312.
 13. Bignardi T, Condous G, Alhamdan D, Kirk E, Van Calster B, Van Huffel S, Timmerman D. The hCG ratio can predict the ultimate viability of the intrauterine pregnancies of uncertain viability in the pregnancy of unknown location population. *Hum Reprod*, 2008 Sep;23(9):1964-7.
 14. Bignardi T, Condous G, Alhamdan D, Kirk E, Van Calster B, Van Huffel S, Timmerman D. The hCG ratio can predict the ultimate viability of the intrauterine pregnancies of uncertain viability in the pregnancy of unknown location population. *Hum Reprod*, 2008 Sep;23(9):1964-7.
 15. Lautmann K, Cordina M, Elson J, Johns J, Schramm-Gajraj K, Ross JA. Clinical use of a model to predict the viability of early intrauterine pregnancies when no embryo is visible on ultrasound. *Hum Reprod*, 2011 Nov; 26(11): 2957 - 63.
 16. Westin M, Källén K, Saltvedt S, Almström H, Grunewald C, Valentin L. Miscarriage after a normal scan at 12-14 gestational weeks in women at low risk of carrying a fetus with chromosomal anomaly according to nuchal translucency screening. *Ultrasound Obstet Gynecol*, 2007 Oct;30(5):728-36

17. Bottomley C, Van Belle V, Kirk E, Van Huffel S, Timmerman D, Bourne T.
Accurate prediction of pregnancy viability by means of a simple scoring system.
Hum Reprod, 2013 Jan;28(1):68-76

FIGURES:

Figure 1:

FLOWCHART SHOWING DATA COLLECTION



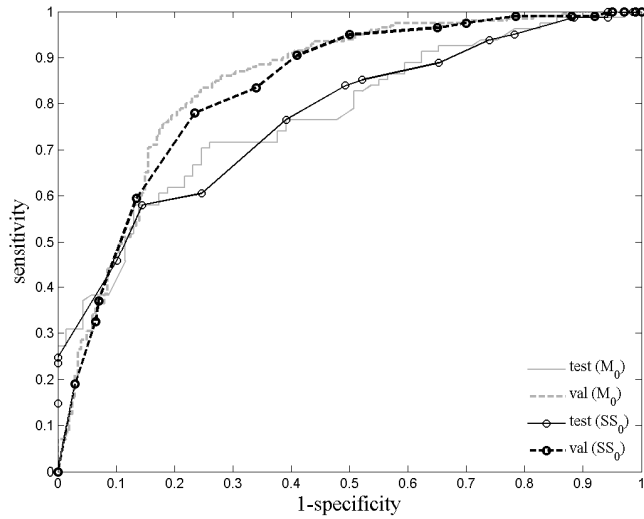


Figure 2: ROC curves for the original model (M_0) and scoring system (SS_n) on the test set of the original paper and the validation set reported in this paper. These models are validated on the set of 400 pregnancies with known outcome and all covariates (including known LMP). Both the model (M_0) and the scoring system (SS_n) perform well on the independent validation data.

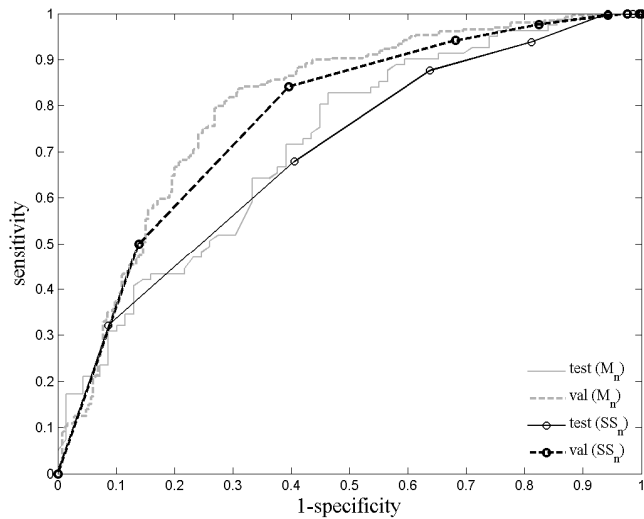


Figure 3: ROC curves for the new model (M_n) and scoring system (SS_n) (not including LMP) on the test set of the original paper and the validation set reported in this paper. These models are validated on the 503 pregnancies with known outcome and all covariates, except LMP. Both the model (M_n) and the scoring system (SS_n) perform well on the independent validation data. However, due to the lack of information on LMP, the performance is less than for the original model (M_0).

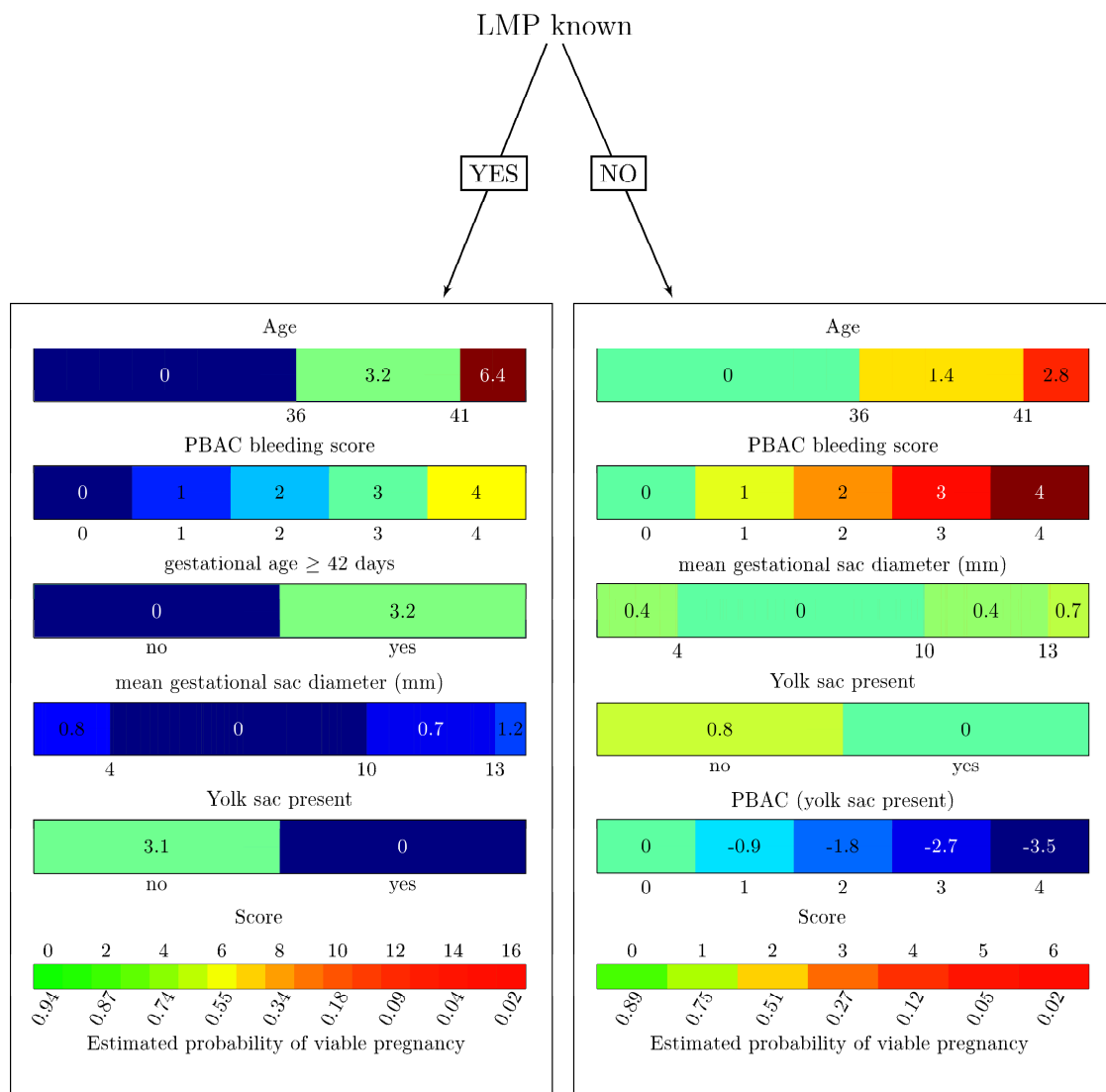


Table 1: Illustration of the new scoring system (SS_n). All points corresponding to the pregnancy characteristics need to be added. The resulting sum is used to obtain an estimated chance on a viable pregnancy by means of the bottom part of the table.

Variable	Points
<i>Age</i>	
≤ 35	0
36- 40	1.4
>40	2.8
<i>PBAC bleeding score</i>	
0	0
1	1
2	2
3	3
4	4
<i>Mean gestational sac diameter</i>	
<4 mm	0.4
4 mm -6.9 mm	0
7 mm - 9.9 mm	0
10 mm -12.9 mm	0.4
≥13mm	0.7
<i>Yolk sac</i>	
present	0
absent	0.8
<i>PBAC bleeding score when yolk sac is present</i>	
0	0
1	-0.9
2	-1.8
3	-2.7
4	-3.5
Estimated probability of viable pregnancy at the end of the first trimester	
0.8900	0
0.7500	1
0.5100	2
0.2700	3
0.1200	4
0.0500	5
0.0200	6